

PNA-Mediated Whiplash PCR

John A. Rose¹, Russell J. Deaton², Masami Hagiya³, and Akira Suyama⁴

¹ Institute of Physics, The University of Tokyo, johnrose@genta.c.u-tokyo.ac.jp

² Department of Computer Science and Computer Engineering,
The University of Arkansas, rdeaton@uark.edu

³ Department of Computer Science, The University of Tokyo,
hagiya@is.s.u-tokyo.ac.jp

⁴ Institute of Physics, The University of Tokyo, suyama@dna.c.u-tokyo.ac.jp

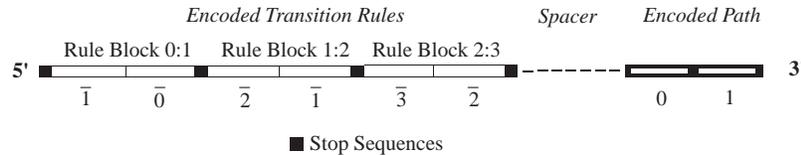
Abstract. In Whiplash PCR (WPCR), autonomous molecular computation is achieved by the recursive, self-directed polymerase extension of a mixture of DNA hairpins. A barrier confronting efficient implementation, however, is a systematic tendency for encoded molecules towards backhybridization, a simple form of self-inhibition. In order to examine this effect, the length distribution of extended strands over the course of the reaction is examined by modeling the process of recursive extension as a Markov chain. The extension efficiency per polymerase encounter of WPCR is then discussed within the framework of a statistical thermodynamic model. The efficiency predicted by this model is consistent with the premature halting of computation reported in a recent *in vitro* WPCR implementation. The predicted scaling behavior also indicates that completion times are long enough to render WPCR-based massive parallelism infeasible. A modified architecture, PNA-mediated WPCR (PWPCR) is then proposed in which the formation of backhybridized structures is inhibited by targeted PNA₂/DNA triplex formation. The efficiency of PWPCR is discussed, using a modified form of the model developed for WPCR. Application of PWPCR is predicted to result in an increase in computational efficiency sufficient to allow the implementation of autonomous molecular computation on a massive scale.

1 Introduction

In Whiplash PCR (WPCR), autonomous computation is implemented by the recursive polymerase extension of a mixture of DNA hairpins [1]. Although the basic feasibility of WPCR has been experimentally demonstrated [1–3], a barrier which confronts efficient implementation is a tendency for single-stranded (ss) DNAs to participate in a form of self-inhibition known as *backhybridization* [1, 2]. To illustrate, consider the WPCR implementation of the 3 step path, $0 \rightarrow 1 \rightarrow 2 \rightarrow 3$, shown in Fig. 1. Computational states are represented by unique DNA words of length, l bases. Each strand is composed of 3 regions. The *transition rule region* encodes the computation’s transition rules (in Fig. 1, $0 \rightarrow 1$, $1 \rightarrow 2$, and $2 \rightarrow 3$). The *head region* contains a record of the strand’s computation, where the 5’-most and 3’-most code words encode for the strand’s initial and current state,

respectively (in Fig. 1, 0 and 1). The *spacer region* guarantees adequate spacing for hybridization. A single round of computation is achieved by the hybridization of the 3' head with a matching code word in the transition rule region, followed by extension by DNA polymerase. Extension is terminated by a short poly-Adenine *stop* sequence, combined with the absence of free dTTP in the buffer. In Fig. 1 (top structure) this process has appended codeword 1 to the strand's 3' end, implementing the transition, $0 \rightarrow 1$. Although the second extension requires the formation of hairpin (a), this process is complicated by the ability of the strand to form *backhybridized* hairpin (b), which is much more energetically favorable than hairpin (a). The number of alternative, backhybridized configurations increases with each extension. For a ssDNA undergoing the r^{th} extension, a total of r alternative hairpin structures will be accessible, only one of which is extendable by DNA polymerase. Occupancy of the $r - 1$ backhybridized structures reduces the concentration of ssDNAs available for the computation.

WPCR Molecule after 1 Successful Extension:



In round 2, a pair of configurations are accessible:

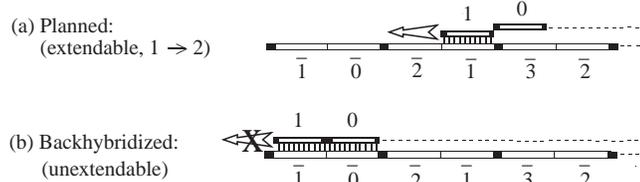


Fig. 1. Backhybridization. After the first extension process (top structure), two hairpins are accessible to the extended molecule. Occupancy of hairpin (b) reduces the concentration of extendable structures (a), and inhibits further computation. A total of $r - 1$ backhybridized structures will be accessible during extension process, r .

In Sec. 2, the length distribution of extended strands, as a function of the reaction temperature and the number of polymerase encounters per strand, is examined by modeling the recursive extension of each strand as a Markov chain. The extension efficiency per polymerase-strand encounter is then discussed using a statistical thermodynamic model of DNA hybridization. Model predictions are shown to be consistent with the premature halting of computation observed in a recent *in vitro* WPCR implementation [3]. Based on the scaling behavior of the model, completion times are predicted to be long enough to render WPCR-based

massive parallelism infeasible. In Sec. 3, a modified architecture, PNA-mediated WPCR (PWPCR) is proposed in which the formation of backhybridized structures is inhibited by targeted PNA₂/DNA triplex formation. The efficiency of PWPCR is then discussed by application of the statistical thermodynamic model developed for WPCR, combined with a simplified all-or-none model of iterative extension. Targeted triplex formation is predicted to be accompanied by a large increase in efficiency, which is sufficient to support the implementation of autonomous molecular computation on a massive scale.

2 The Efficiency of Whiplash PCR

The appeal of WPCR lies in the potential for the parallel implementation of a massive number of distinct computational paths. For this purpose, a distinct DNA species must be included in the initial reaction mixture for each acyclic path in the instance graph. Although a general analysis of hairpin extension efficiency would require an assessment of strand-strand interaction, in WPCR the DNA molecules are anchored to a solid support. As a result, the impact of intermolecular interaction may be neglected, allowing the recursive extension of each WPCR species to be modeled independently. The fundamental details of WPCR efficiency are therefore contained in an analysis of the single-path case.

The process of recursive extension for each DNA strand may be modeled as a Markov chain [4]. For a q -step WPCR implementation, let the *extension state*, r of each strand be defined to equal the number of times the molecule has been successfully extended plus 1. Note that a strand's extension state is distinct from a strand's *computational state*. During the course of the reaction, extending strands may occupy a total of $q + 1$ extension states, ranging from $r = 1$ (completely unextended) to $r = q + 1$ (fully extended). Let ϵ_r denote the probability that a polymerase encounter with a DNA strand in extension state r observes the strand in an extendable configuration. With each polymerase encounter, a DNA strand will increment its extension state by either 0 or 1, with probabilities $1 - \epsilon_r$, and ϵ_r , respectively. For molecules which reach the final absorbing state, $q + 1$, no further extension is possible (*i.e.*, $\epsilon_{q+1} = 0$). The state occupancies resulting from N_e polymerase encounters/strand at temperature T_{rx} are given by the product of the N_e -step transition matrix, $\mathbf{P}(T_{rx}, N_e)$ and the initial state occupancy vector, $[N_o \ 0 \ \dots \ 0]$, where N_o is the total strand number. $\mathbf{P}(T_{rx}, N_e)$ is given by the Chapman-Kolmogorov eq. [4],

$$\mathbf{P}(T_{rx}, N_e) = \begin{bmatrix} 1 - \epsilon_1 & \epsilon_1 & \dots & 0 & 0 \\ 0 & 1 - \epsilon_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 - \epsilon_q & \epsilon_q \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}^{N_e}. \quad (1)$$

The estimation of N_e and ϵ_r is discussed in Sec. 2.1 and Sec. 2.2, respectively. The resulting state occupancies estimate the length distribution, in terms of

number of extensions, among all N_o strands, for particular values of T_{rx} and N_e . Accounting for a more complicated thermal program is straightforward. For a thermal cycle which consists of several polymerization periods of diverse duration and temperature, the process of extension is modeled by (1) estimating an N_e value for each subcycle, (2) constructing a transition matrix for each subcycle according to the T_{rx} employed, and (3) applying the resulting set of matrices iteratively to the initial state occupancy vector.

2.1 The Efficiency per Polymerase-DNA Encounter

The quantity ϵ_r may be discussed within the framework of a statistical thermodynamic model. Consider an ensemble, S_r of identical WPCR molecules, each of which has been extended $r-1$ times. Assuming an all-or-none model of duplex formation, members of S_r will be distributed amongst $r+1$ configurations: an unfolded ssDNA species, an extendable hairpin species, and a set of $r-1$ unextendable hairpin species, each of which is a backhybridized artifact from a previous round of extension. The statistical weight of a simple hairpin configuration, which consists of an end loop of n unpaired bases and a lone duplex of length j paired bases is estimated by $K = \sigma Z_j (n+1)^{-1.5}$, where Z_j is the statistical weight of stacking and σ is the cooperativity parameter [5].

In order to ensure the uniformity of the various extension reactions of an implementation, WPCR code words are typically selected to have uniform GC content [2]. This procedure results in an approximately equal Gibbs free energy of stacking for each codeword with its Watson-Crick complement [3]. The statistical weight of stacking for a length j duplex is then estimated by $Z_j = s^{j-1}$ [6], where s is the statistical weight for the average base pair doublet of the implementation. The equilibrium fraction of extendable ensemble members, ϵ_r is estimated by the ratio of the statistical weight of the extendable hairpin to the sum of the statistical weights of all structures. Constructing this ratio with the particular values, $j = l$ and $j = 2l$ for the single planned, and $r-1$ backhybridized hairpin configurations, respectively yields,

$$\epsilon_r = \left[1 + \gamma_r s^l + \frac{(n_r + 1)^{1.5}}{\sigma s^{l-1}} \right]^{-1}, \quad (2)$$

for the extension efficiency per polymerase-DNA encounter of the single-path WPCR implementation. Here, $\gamma_r \approx \sum_{i=1}^{r-1} (n_r/n_i)^{1.5}$ expresses the impact of variations in loop length between competing hairpin structures, n_r is the terminal loop length of the extendable configuration, and each n_i is the loop length of the hairpin structure extended during previous round i .

The single path case may be generalized to apply to parallel WPCR if variations in ϵ_r due to differences in the specific ordering of transition rule blocks within the rule region are neglected. It is straightforward to demonstrate that the values, $\bar{\gamma}_r \approx 1.66r$ and $\bar{\pi}_r \approx (q+r)l$ are those characteristic of an implementation with mean loop lengths in all rounds, where the average is taken over all

transition rule orderings. Combining these mean values with Eq. 2 yields,

$$\bar{\epsilon}_r \approx \left\{ 1 + 1.66rs^l + \frac{[l(q+r)]^{1.5}}{\sigma s^{l-1}} \right\}^{-1} \quad (3)$$

for the mean efficiency of a parallel, q -step WPCR implementation with parameters l and s . This expression may also be used to estimate the efficiency of the mean q -step single-path implementation. In the following text, estimates which have been obtained using $\bar{\epsilon}_r$ will be distinguished by an overscore.

2.2 The Mean Polymerase/DNA Encounter Rate

The mean number of polymerase encounters per strand, during a polymerization period of length Δt_p may be estimated as follows. Let N_u denote the number of units of *Taq* DNA polymerase utilized, where 1 unit corresponds to the synthesis of 10 nmol of added bases in 30 minutes, using an excess of activated salmon sperm DNA as substrate [7]. Let v_t denote the number of distinct extensions/second by 1 unit of polymerase under optimal conditions, using excess substrate (target and primer), and in the absence of unextendable substrate. *Taq* DNA polymerase is fast and highly processive [7]. It is therefore assumed that (1) the mean polymerase-DNA dissociation time is much larger than both the time required for oligo-length extension and the mean time between polymerase-DNA encounters, and (2) each encounter results in the all-or-none (oligonucleotide length) extension of the encountered molecule. In this case, the total number of enzyme-substrate encounters in time Δt_p is invariant to the DNA substrate extendability, and may be estimated by the product $N_{enc} = N_u v_t \Delta t_p$. Assuming that encounters are distributed uniformly over all N_o strands, the number of encounters/strand which occur in time Δt_p is estimated by,

$$N_e = \frac{N_{enc}}{N_o} = \frac{N_u v_t \Delta t_p}{N_o}. \quad (4)$$

2.3 Comparison with Experiment

The WPCR implementation of an 8 step path was recently reported [3]. The experimental protocol in [3] was as follows. An estimated total of $N_o \approx 1.2 \times 10^{13}$ immobilized strands was utilized, with 5 units of *Taq* DNA polymerase, in a total volume of 400 μ L. Constant conditions of pH = 7.0 and I = 0.205 M ($[K^+] = 0.05$ M, $[Mg^{++}] = 1.5$ mM) were maintained. The first extension process of each strand was implemented separately, by ‘‘input PCR’’. The remaining 7 extensions were implemented by the application of 15 thermal cycles, each of which consisted of (1) 30 s at 337 K, (2) a rapid increase to 353 K in 60 s, (3) 300 s at 353 K, and (4) a decrease to 337 K in 120 s. The success of each extension was evaluated in all-or-none fashion, by means of a novel ‘‘output PCR’’ technique. Success of the output phase was evaluated using gel electrophoresis. Bright bands were reported at the mobilities characteristic of the fully extended product for

each of the first 5 extensions (including the extension implemented by input PCR). This result was taken to indicate the success of the first five extensions. Very faint bands reported at various other mobilities are assumed to indicate error extension during WPCR and output PCR.

In [3], it was maintained that problems due to backhybridization had been overcome by the applied thermal program, and that the observed poor performance was due to other factors. The validity of this view may be tested theoretically by a comparison of the observations reported in [3] with the predictions of the Markov chain model. For this purpose, the free energies of the code word set in [3] were estimated using the nearest-neighbor model of [8]. Computed values were verified to approximately satisfy the assumption of code word energetic uniformity. For instance, the mean code word standard enthalpy and entropy of stacking for each $l = 15$ base DNA code word was estimated at 114 ± 2.04 kcal/mol and 303 ± 5.62 cal/mol K, respectively, at 1.0 M $[\text{Na}^+]$. Values were then adjusted to account for the reported experimental K^+ and Mg^{++} concentrations, using the methodology described in [9]. The statistical weight of the mean single stacked doublet in [3] was then estimated from the Gibbs free energy of stacking, $\langle \Delta G^\circ \rangle$ by the Gibbs factor, $s_{nn} = -\langle \Delta G^\circ \rangle / RT_{rx}$, where R is the ideal gas constant. The consensus value of the cooperativity parameter, $\sigma = 4.5 \times 10^{-5}$ was assumed [6]. The temperature dependence of $\bar{\epsilon}_r$ was estimated for the implementation in [3] using Eq. 3. A maximal extension efficiency per encounter of roughly 3×10^{-5} is predicted at 350 K. This predicted optimal T_{rx} is in good agreement with the experimentally determined optimum of 353 K.

In addition to the parameters discussed above, an estimation of overall efficiency requires an estimate of v_t . The estimate, $v_t \approx 6.70 \times 10^{10}$ encounters/unit/s, was obtained by taking the ratio of the rate of nucleotide addition defined to equal 1 unit of enzyme, and the mean number of bases added per polymerase-DNA encounter. Based on the manufacturer's estimate, a mean processivity of 50 bases/encounter was assumed [11]. The present Markov chain model of recursive extension, has been used to estimate the number of strands, \bar{N}_r in [3] having undergone each of from 1 ($r = 2$) to 8 ($r = 9$) extensions, as a function of thermal cycle. Results are illustrated in Fig. 2(a). The implementation of the first extension by input PCR was modeled by assigning an efficiency of unity for the first extension. As shown, the production of fractions of molecules which have successfully undergone from 1 ($r = 2$) to 4 ($r = 5$) extensions is predicted during the first thermal cycle. The production of longer strands, however, is delayed until the 11th cycle, when the appearance of 5-fold extended ($r = 6$) molecules is predicted. The production of 6 to 8-fold extended ($r > 6$) molecules is not predicted to occur during the course of the experiment. These predictions are in agreement with the experimental behavior reported in [3], which reported the production of strands with up to 5 extensions. This agreement between model predictions and experimentally observed behavior lends strong support to the theory that backhybridization was responsible for the premature failure observed in [3], and calls into question the success of the isothermal protocol in eliminating problems stemming from backhybridization.

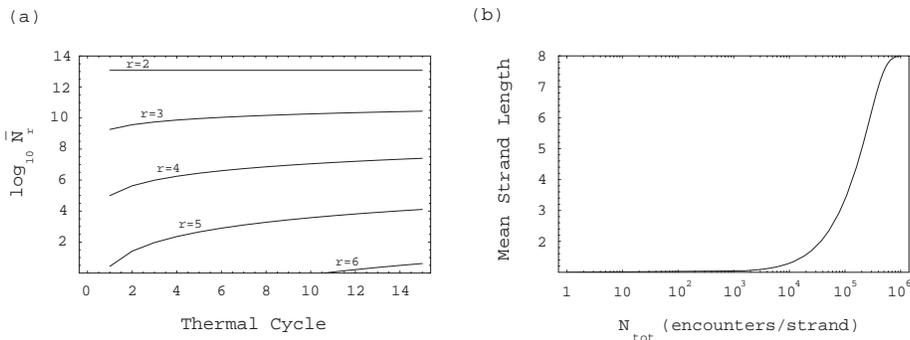


Fig. 2. The Efficiency of WPCR. (a) The mean number of strands, \overline{N}_r , predicted to undergo a total of from 1 extension ($r = 2$) to 5 extensions ($r = 6$), as a function of thermal cycle, for the WPCR implementation in [3]. The total strand number was roughly 1.2×10^{13} . (b) Mean strand length, in terms of extension number, as a function of the total number of polymerase encounters/strand, N_{tot} .

Continued application of a large number of thermal cycles must eventually result in completion. However, this process is predicted to require unrealistic reaction time. As shown in Fig. 2(b), the WPCR implementation in [3] (adjusted to the optimal $T_{rx} = 350$ K) is predicted to require $\approx 5 \times 10^4$ polymerase encounters/strand to exceed a mean efficiency of 2 extensions/strand. At the estimated rate of 8.4 encounters/strand/5 minute round, this corresponds to a total time of ≈ 500 hours. Furthermore, 4.0×10^5 encounters/strand are required to reach a mean of 7 encounters/strand (165 days). Mean completion is reached at roughly 10^6 encounters/strand (1.1 years). The linear scaling of encounter number predicted with N_u (*cf.*, Eq 4) also indicates that an attempt to reduce reaction time by using excess polymerase will encounter limited success. For instance, if $N_u = 54$ units of polymerase are used (90.7 encounters/round), the completion time for the 8-step path in [3] is reduced to 38 days.

3 PNA-mediated WPCR

3.1 Inhibiting Backhybridization

WPCR may be redesigned to enable the specific inhibition of backhybridized structures by targeted PNA₂/DNA triplex formation. The ability of peptide nucleic acid strands (PNAs) to bind to complementary ssDNA with extremely high affinity and sequence-specificity is well characterized [12]. For a pair of homopyrimidine PNA strands, binding to a complementary ssDNA target sequence occurs with stoichiometry 2 PNA:1 DNA, indicating the formation of a PNA₂/DNA triplex. Under appropriate reaction conditions, rapid, irreversible formation of the triplex structure occurs, even if the target sequence is embedded in a dsDNA duplex. This strand invasion results in the extrusion of the target-complementary DNA strand, forming a “P-loop” [13].

The rule block structure of WPCR may be modified to enable directed triplex formation. In particular, separation of each source/target codeword pair by the sequence, $T_4CT_2CT_2$ results in the separation of state-encoding sequences in the head region by $A_2GA_2GA_4$, the target sequence of the highly efficient cationic bis-PNA molecule reported in [14]. This is shown in Fig. 3(a). Exposure of the reaction mixture, after each polymerization round to a low $[Na^+]$, excess [bis-PNA] wash then results in a high saturation of target sequences with bis-PNA (Fig. 3, panel b). For the reported first-order rate constant of 2.33 min^{-1} at $1.0 \mu\text{M}$ bis-PNA, $20 \text{ mM } [Na^+]$ [14], a fractional saturation of 0.999 is achieved within 3 min. Cytosine-bearing, cationic bis-PNAs of length 10 bases have been reported to melt from complexed ssDNA at $\approx 85^\circ \text{ C}$ (in $0.1 \text{ M } [Na^+]$), with a very narrow melting transition [15]. The maintenance of PNA_2/DNA triplexes formed during the bis-PNA wash, during subsequent polymerization can therefore be assured by the selection of a polymerization temperature substantially less than 80° C . In each round, the presence of a PNA_2/DNA triplex immediately 5' to the new head region will not inhibit planned hybridization, due to the extreme compactness of the P-loop. The stability of the extended backhybridized configuration (shown in Fig. 3, structure c1), however will be diminished due to the separation of the duplex islands by a PNA_2/DNA triplex. This modified protocol will be referred to as PNA-mediated WPCR (PWPCR).

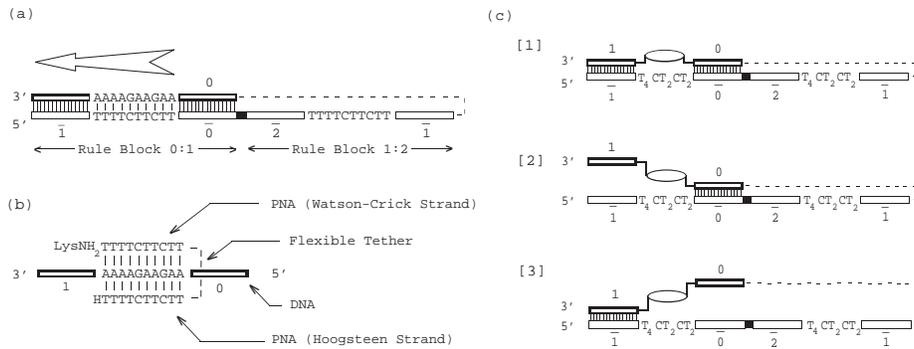


Fig. 3. PNA-mediated WPCR. [a] A “target” sequence, $A_2GA_2GA_4$ is produced between codewords during each extension. [b] Addition of bis-PNA results in the formation of a PNA_2/DNA triplex at the target sequence. Triplex represented by an oval in subsequent structures. [c] Accessible backhybridized structures have decreased stability relative to those in WPCR.

3.2 The Efficiency of PNA Mediated WPCR

The effect of the presence of the PNA_2/DNA triplex on the stability of hybridized structures, and on the per encounter polymerization extension efficiency may be

estimated by means of a statistical thermodynamic model. Due to the experimentally reported compactness of the P-loop, the presence of a triplex region immediately adjacent to the head sequence is assumed to have a negligible impact on the ability of the head to hybridize with a complementary sequence in the transition rule region. Each successful extension may facilitate the later formation of three distinct backhybridized structures (see Fig. 3): (1) an extended structure, composed of a pair length l duplex islands punctuated by a P-loop (structure C1), or (2,3) two shorter structures, each of which is generated by formation of one of the duplex islands of the extended structure (structures C2, C3). Like the planned configuration, backhybridized hairpins C2 and C3 each have the form of a simple hairpin structure, with a statistical weight given by $K = \sigma Z_{l-1} (n+1)^{-1.5}$, where n is the terminal loop length of the particular structure, as discussed in Sec. 2.1. Here, the longer (n'') and shorter (n') of the associated terminal loop lengths are related by $n'' = n' + 10 l/3$. The statistical weight of the extended backhybridized configuration, C1 has the form $Z_i = Z_p \sigma^2 s^{2l-2} (1+n'_i)^{-1.5}$, where Z_p is the post-triplex formation statistical weight of the P-loop. As discussed earlier, given the use of a polymerization temperature substantially less than 80° C, the presence of the triplex may be assumed (statistical weight of 1). Z_d then reduces to the statistical weight of interaction between the P-loop components (*i.e.*, the established triplex and the extruded single strand). The P-loop's distinctive eye structure [13] suggests the absence of stabilizing interactions between the extruded single-stranded target-complementary strand and the PNA₂/DNA triplex. Z_p is therefore assumed to be entirely entropic in origin, and is modeled as a Gaussian chain with excluded volume. For a target region of length $\frac{2}{3} l$, the loop region is assigned a statistical weight of $Z_p = (2 + 4l/3)^{-1.7}$. Taking the ratio of the statistical weight of the expected configuration to that of all configurations, and assuming the mean transition rule ordering, yields

$$\bar{\epsilon}'_r \approx \left\{ 1 + 1.94r \left[2 + \frac{\sigma s^{l-1}}{(2 + 4l/3)^{1.7}} \right] + \frac{[l(2r + 4q/3)]^{1.5}}{\sigma s^{l-1}} \right\}^{-1}, \quad (5)$$

for the extension efficiency/polymerase-substrate encounter for a strand undergoing the r^{th} extension process, in the mean-path PWPCR implementation with static characteristics l , q , and s . A comparison of expressions 3 and 5 indicates that the primary effect of targeted PNA₂/DNA triplex formation on the per encounter extension efficiency is the destabilization of the full length backhybridized configuration by a factor of σ .

3.3 The Overall Extension Efficiency

The Markov chain model of extension used to discuss WPCR may also be applied to PWPCR. This procedure, however is complicated by the need to separate the PNA treatment from each extension process. In particular, the two processes may not be performed concurrently, because of the very low ionic strength required for high efficiency PNA₂/DNA triplex formation. As a result,

application of a Markov chain model requires the definition of an additional set of intermediate states, and the use of a second transition matrix, to model the formation of triplexes during each PNA treatment. A simpler stochastic model of performance, however may be constructed by modeling the extension process for each strand, during the polymerization period of each PWPCR cycle as a single-step, all-or-none transition. This approximate treatment, which has the advantage of yielding a closed form estimate of completion efficiency, is motivated by the extremely low efficiency per polymerase encounter predicted for a P-WPCR molecule which has been extended but not treated with bis-PNA, due to the increased length of the non-PNA treated backhybridized structure.

Consider the observation of a ssDNA which has been successfully extended in each of a total of $c - 1$ PWPCR cycles. The probability that all of the N_e polymerase encounters with this molecule that occur in the polymerization period of cycle c will result in extension failure is equal to $(1 - \bar{\epsilon}'_c)^{N_e}$. The probability of successful extension is then estimated by, $p_{ext}(c) = 1 - (1 - \bar{\epsilon}'_c)^{N_e} \approx N_e \bar{\epsilon}'_c$. If $\langle N_{c-1} \rangle$ denotes the mean number of fully extended DNA strands present in a WPCR mixture at the end of cycle $c - 1$, then the mean number of fully extended structures present after cycle c can be written as $\langle N_c \rangle = \langle N_{c-1} \rangle p_{ext}(c)$. This relationship may be applied $c - 1$ times to yield the estimate,

$$\chi(c) \equiv \frac{\langle N_c \rangle}{N_o} \approx \frac{N_e^{c-1}}{N_o} \prod_{i=2}^c \epsilon'_i, \quad (6)$$

for the fraction of c -fold extended strands produced after cycle c . Here, the first extension process for each strand in the first cycle has been assumed to proceed with an efficiency of unity, due to the absence of backhybridization.

The impact of PNA₂/DNA triplex formation on the overall efficiency of computation may be illustrated by concrete application. For this purpose, the efficiency of a PWPCR implementation of the 8-step computational path described in [3], in terms of the log of the number of fully extended substrate molecules, where $\langle \bar{N}_c \rangle = N_o \bar{\chi}(c)$, was estimated using Eqs. 5 and 6, and is illustrated in Fig. 4(a). For consistency, a codeword set energetically equivalent to the set presented in [3] was assumed. Buffer conditions and total polymerization time were also assumed to be identical to [3]. A comparison of Fig. 2(a) and Fig. 4(a) indicates that the triplex-induced inhibition of backhybridization results in a substantial increase in predicted overall efficiency of computation. If each extension is performed at the predicted optimal reaction temperature of 60° C, roughly 1.3×10^9 of the initial 1.2×10^{13} encoded strands are predicted to be fully extended after the completion of all rounds.

3.4 The Parallelization of PWPCR

The ultimate aim of both WPCR and PWPCR is to effect the parallel, *in vitro* simulation of a massive number of distinct paths. Consider the parallelization of a PWPCR implementation, in which the set of N_o strands has been parsed into P distinct species, each of which represents a different computational path, and

is present with equal copy number, $N_{copy} = N_o/P$. The maximum parallelism obtainable by this implementation is equal to $P = N_o/N_{copy}$. However, this is practically obtained only when N_{copy} is sufficiently large to ensure the full extension of at least one copy per path. Given the absence of bimolecular interaction, it is straightforward to demonstrate that the threshold of completion for a parallel PWPCR implementation is reached when N_{copy} is chosen such that,

$$[1 - \bar{\chi}'(q)]^{N_{copy}} = \frac{N_{copy}}{N_o}. \quad (7)$$

For the 8-step PWPCR implementation under discussion, $N_o = 1.2 \times 10^{13}$ and $\bar{\chi}'(8) = 1.1 \times 10^{-4}$. According to Eq. 7, maximum parallelism for this implementation is achieved when $N_{copy} \approx 1.6 \times 10^5$. This copy number yields a maximum parallelism of $P \approx 7.5 \times 10^7$, which corresponds to the implementation of roughly $N_{ops} = qP \approx 1.5 \times 10^9$ distinct computational operations. A substantial improve-

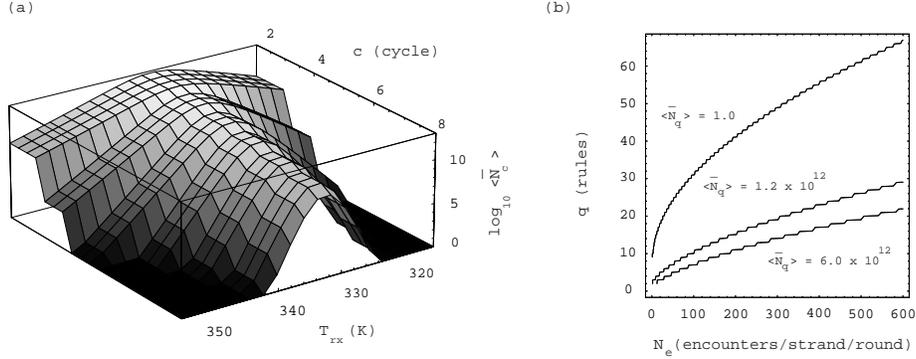


Fig. 4. The Efficiency of PWPCR (a) An estimate of the number of c -fold extended strands, $\langle \bar{N}_c \rangle$ as a function of T_{rx} , c PWPCR cycles, for the mean implementation of an 8 step computational path. (b) The contours which define the line of failure, 10% efficiency, and 50% efficiency after all q cycles ($\langle \bar{N}_q \rangle = 1.0, 1.2 \times 10^{12}$, and 6.0×10^{12} , respectively), vs. N_e and q , for PWPCR implementations of length $q = 2 - 65$ rules. Accompanying surface and z-axis omitted for clarity.

ment may be obtained by modest modification of the protocol. Fig. 4 illustrates the contours which are predicted to define the lines of 0%, 10%, and 50% completion efficiency per strand for PWPCR implementations of length $q = 2 - 65$, as a function of N_e and q . The use of a codeword set energetically equivalent to that in [3] was assumed. Under this protocol, the application of a realistic set of extension reaction conditions ($N_u = 54$ units, $\Delta t_p = 30$ min, yielding $N_e \approx 542.7$) allows the implementation of computational paths of length $q = 20$, with a per strand efficiency at completion of $\bar{\chi}'(20) \approx 0.5$. According to Eq. 7, the maximum parallelism under these conditions, $P \approx 3.1 \times 10^{11}$ paths, is achieved when

$N_{copy} \approx 40$ copies per path. This corresponds to the implementation of $N_{ops} \approx 6.2 \times 10^{12}$ distinct operations.

4 Conclusion

In this work, the impact of backhybridization on WPCR efficiency was investigated by modeling the extension of each hairpin as an independent Markov chain, and estimating the associated state transition probabilities using the statistical thermodynamic theory of DNA melting. This model was shown to predict that the poor performance of WPCR observed in [3] was due to backhybridization. This is significant, because in [3], it was maintained that problems due to backhybridization had been overcome by the applied thermal program, and that the observed poor performance was due to other factors. The scaling behavior of the model also predicts that mean completion times are sufficiently long to render WPCR impractical for massive parallelism. In an effort to enhance computational efficiency by reducing the impact of backhybridization, a modified architecture, PWPCR was then introduced, which enables the specific inhibition of backhybridized structures through targeted PNA₂/DNA triplex formation. Application of this protocol is predicted to result in an efficiency increase which is sufficient to allow the realistic implementation of massive parallelism.

Acknowledgements. The authors are grateful to A. Nishikawa of the Osaka Electro-Communication Junior College and M. Arita of the Tokyo Electrotechnical Laboratory for critical reviews of the manuscript. The authors are also grateful to the referee for helpful comments. Financial support was provided by the JSPS “Research for the Future” Program (JSPS-RFTF 96I00101).

References

1. M. Hagiya *et al.*, in [16], p. 57.
2. K. Sakamoto *et al.*, *Biosystems* **52**, 81 (1999).
3. K. Komiya *et al.* in *DNA Computing*, edited by A. Condon and G. Rozenberg, (Springer-Verlag, Berlin, 2001), p. 19.
4. S. Ross, *Intro. to Probability Models*, 7th Ed., (Academic Press, San Diego, 2000).
5. D. Poland and H. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
6. R. Wartell and A. Benight, *Physics Reports (Review Section of PRL)* **126**, 67 (1985).
7. M. Innis *et al.*, *Proc. Natl. Acad. Sci.* **85**, 9436 (1988).
8. J. SantaLucia, Jr., *Proc. Natl. Acad. Sci.* **95**, 1460 (1998).
9. J. Wetmur, in [16], p. 1.
10. S. Kozyavkin, S. Mirkin, and B. Amirikyan, *J. Biomol. Struct. Dyn.* **5**, 119 (1987).
11. <http://www.pebio.com/pc/catalog/pg9.html>
12. A. Lomakin and M. Frank-Kamenetskii, *J. Mol. Biol.* **276**, 57 (1998).
13. D. Cherny *et al.*, *Proc. Natl. Acad. Sci.* **90**, 1667 (1993).
14. H. Kuhn *et al.*, *Nuc. Acids Res.* **26**, 582 (1998).
15. M. Griffith *et al.*, *J. Am. Chem. Soc.* **117**, 831 (1995).
16. H. Rubin and D. Wood (Eds.), *DNA Based Computers III*, (American Mathematical Society, Providence, RI, 1999).